A Bayesian analysis for fitting manifolds of varying dimensions

Philip Torr
Microsoft Research
http://www.research.microsoft.com

Microsoft Research

- Set up about 6 years ago.
- Redmond 400-500 researchers
- Cambridge 40-50 researchers
- China 100 researchers

Microsoft Research

- Redmond Anandan, Szeliski, Shafer
- Cambridge Bishop, Blake, Torr
- China Shum, Lee

Microsoft Research Beijing

- Shum, graphics, light fields, smart texture generation for games
- Other, face detection, video parsing and understanding.

Microsoft Research Cambridge

- MLP group, headed by Chris Bishop
- Me, matching, 3d reconstruction, some work on face detection, video editing and understanding
- Blake, work on image cut out, tracking.
- Herbrecht, Tipping, game AI

Outline of Talk

- First, review old work (with Fitzgibbon and Zisserman), on the need for model selection in SAM (Structure and Motion recovery) problem.
- Second, examine model selection paradigms for manifold fitting.
- Third, Bayesian analysis.

AIM

- To produce some easily computable bounds on the Bayesian solution (the evidence) without resorting to MCMC.

- Assumptions
  - The error distribution is rotationally symmetric.
  - The manifold has mostly low curvature.

Scenario: Sam
Structure and Motion Recovery

1. Need to recover the matches between the images.

2. Recovering the matches is equivalent to recovery of structure.

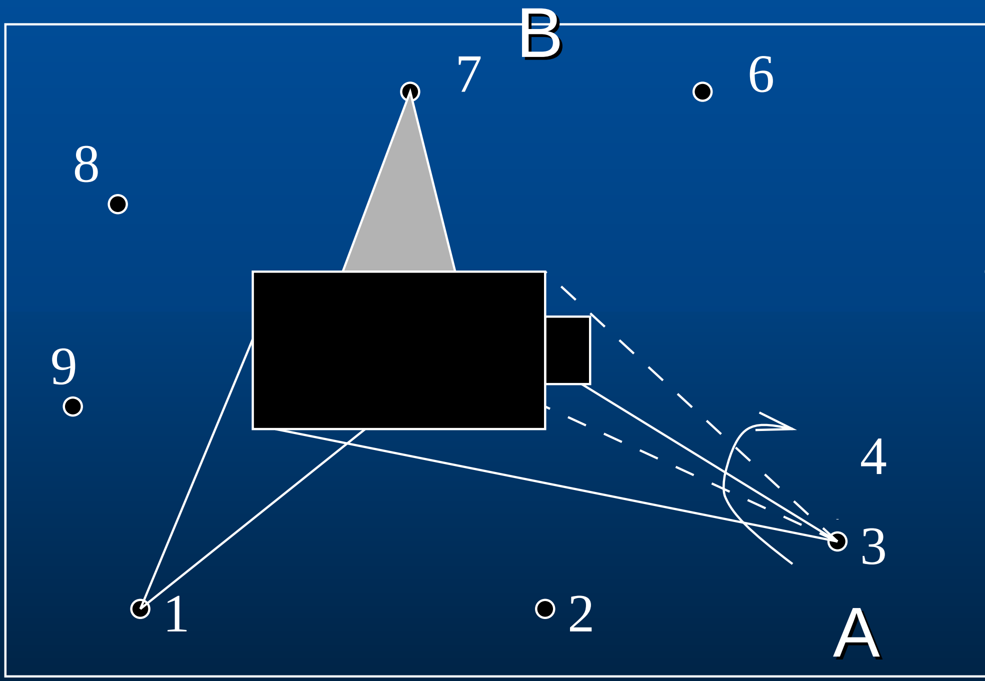
3. To recover matches need to recover the rigidity constraint, to guide matching: epipolar geometry or homography.

Problem Degeneracy F or H?: Torr, Fitzgibbon & Zisserman 97, 99

Two problem cases for SAM (structure and motion) recovery.

A Camera Rotation

B Planar object.



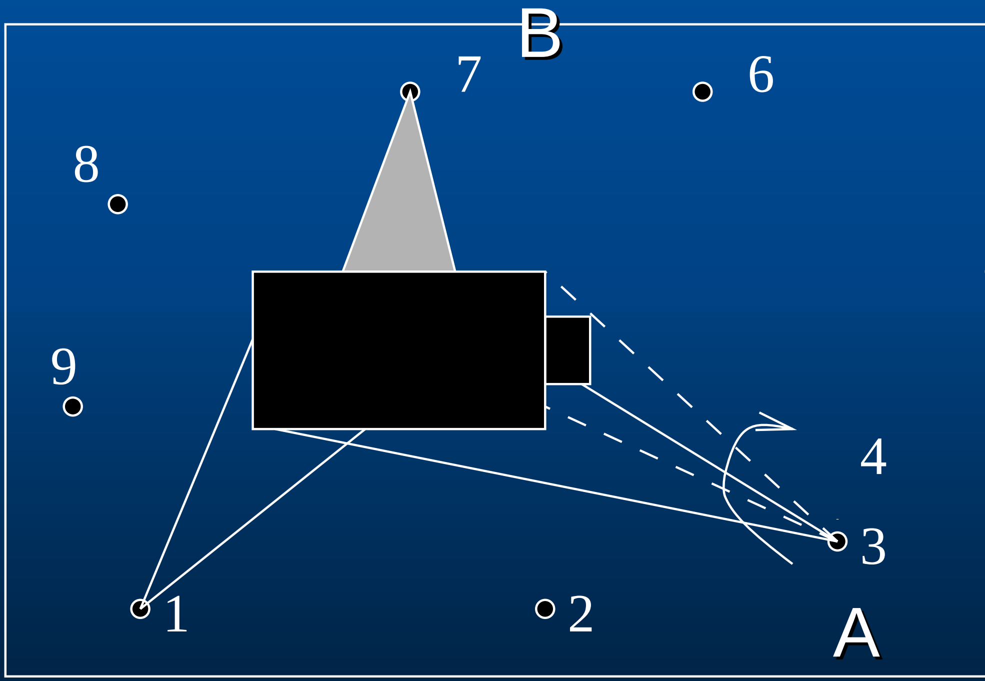
Problem: Model Selection to determine whether **F** or **H**

Problem Degeneracy F or H?: Torr, Fitzgibbon & Zisserman 97, 99

Two problem cases for SAM (structure and motion) recovery.

A Camera Rotation

B Planar object.



Solution: was to use **GRIC** criteria, to be explained later.

When homography describes scene:

- A: Camera rotates, no new structure information.
- B: 2 views have a plane in common; can not put structure into the same projective frame (3 degrees of freedom).
- Note, new work of Pollefreys, Verbiest, Gool…

Pollefreys, Verbiest, Gool
ECCV 2002 to appear

- If **F**(1,2) & **F**(2,3) can be recovered, but all points common to 1 & 3 are on a plane.
- Then there is a one parameter family of projective reconstructions.
- Their solution: is to use GRIC (Torr et al) to detect planes and self calibration to resolve the ambiguity.

This is an example of fitting manifolds of varying dimension:

- 2 Views----Consider Image coordinates 4D space

	Dimension 3	Dimension 2
Bilinear	**F** Matrix	Homography **H**
Linear	Affine **F** Matrix	Affinity **A**

- (non generic: quadratic transformations, dimension 2.)

- Three views: same dimension for manifolds.

Concatenated or Joint Image Space

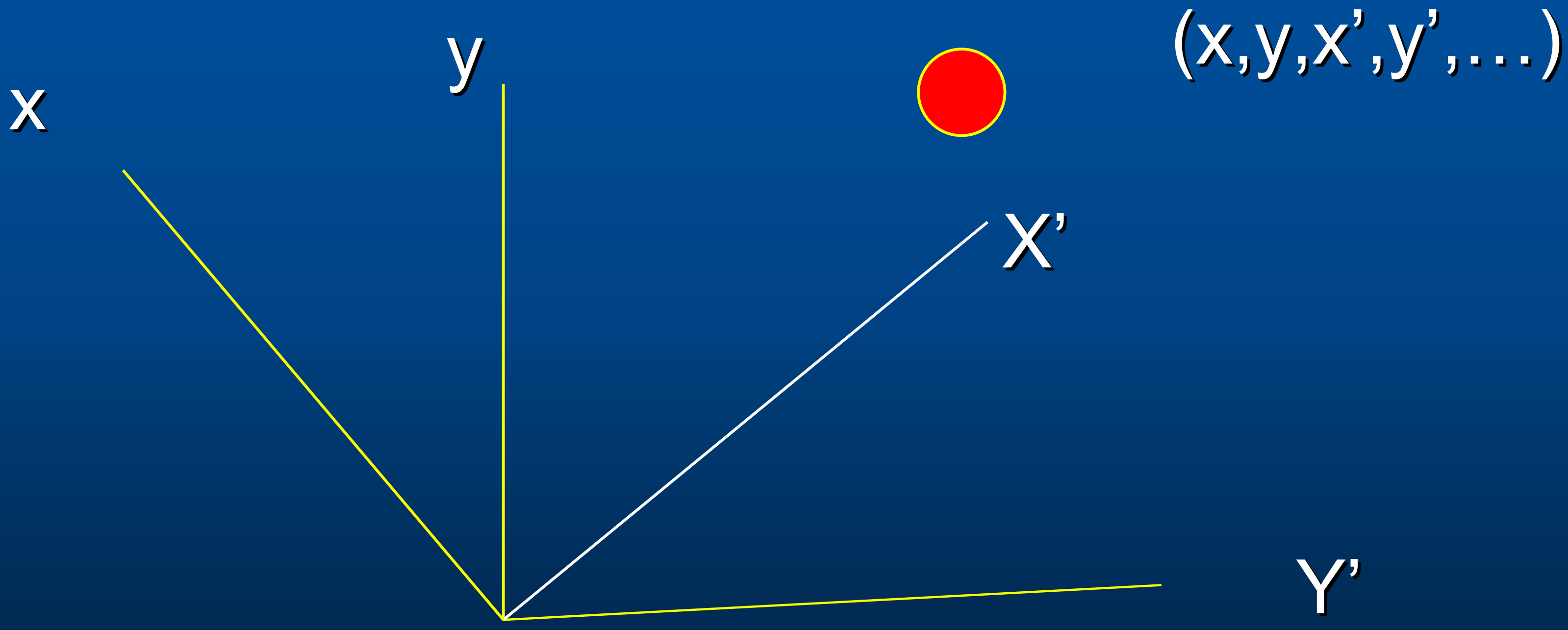
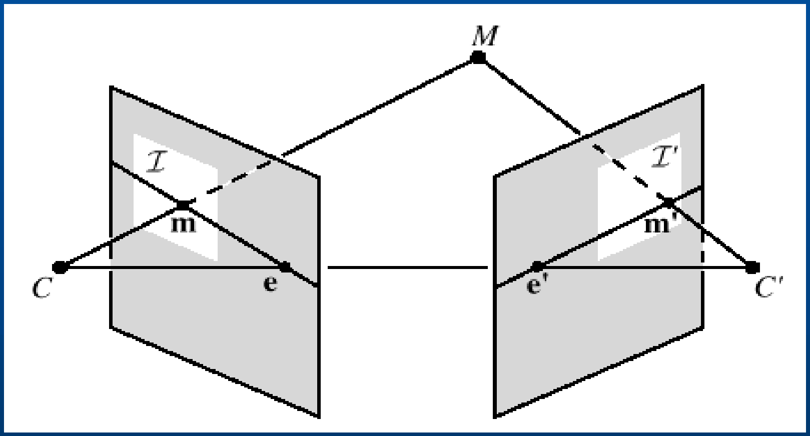
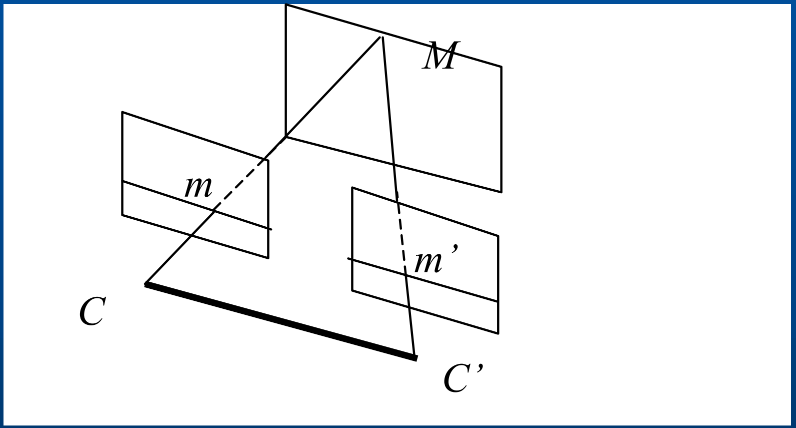


Image coordinates in higher dimensional space

Guide matches with Epipolar Geometry





- Epipoles anywhere
- Fundamental matrix

  F: a 3x3 rank-2 matrix

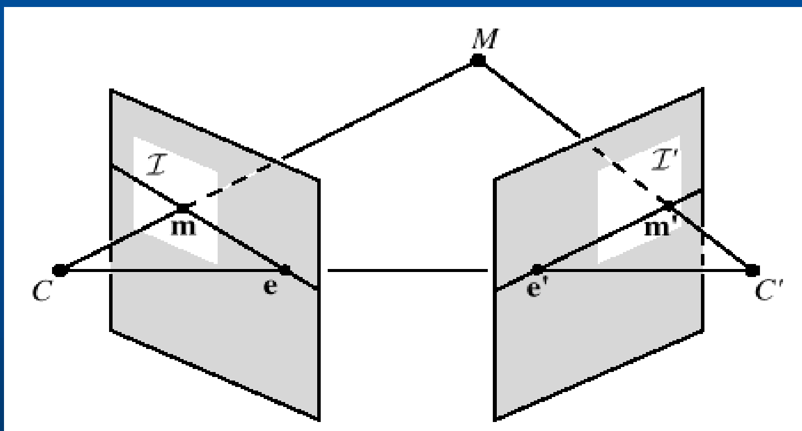
  7 DOF

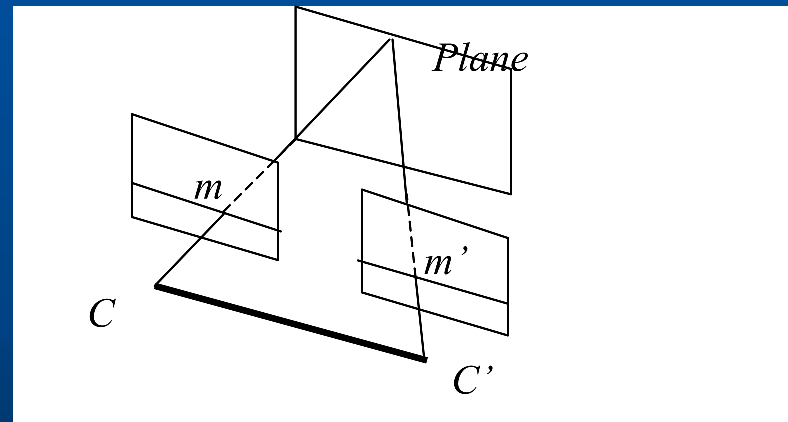
- No Epipole define
- Homography

  H: a 3x3 matrix

  8 DOF

Guide matches with Geometry





$$\mathbf{x}^t \mathbf{F} \mathbf{x}' = 0$$

$$\mathbf{x}' = \mathbf{H}\mathbf{x}$$

$$\begin{pmatrix} x & y & 1 \end{pmatrix} \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = 0$$

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Taxonomy of Motion Models

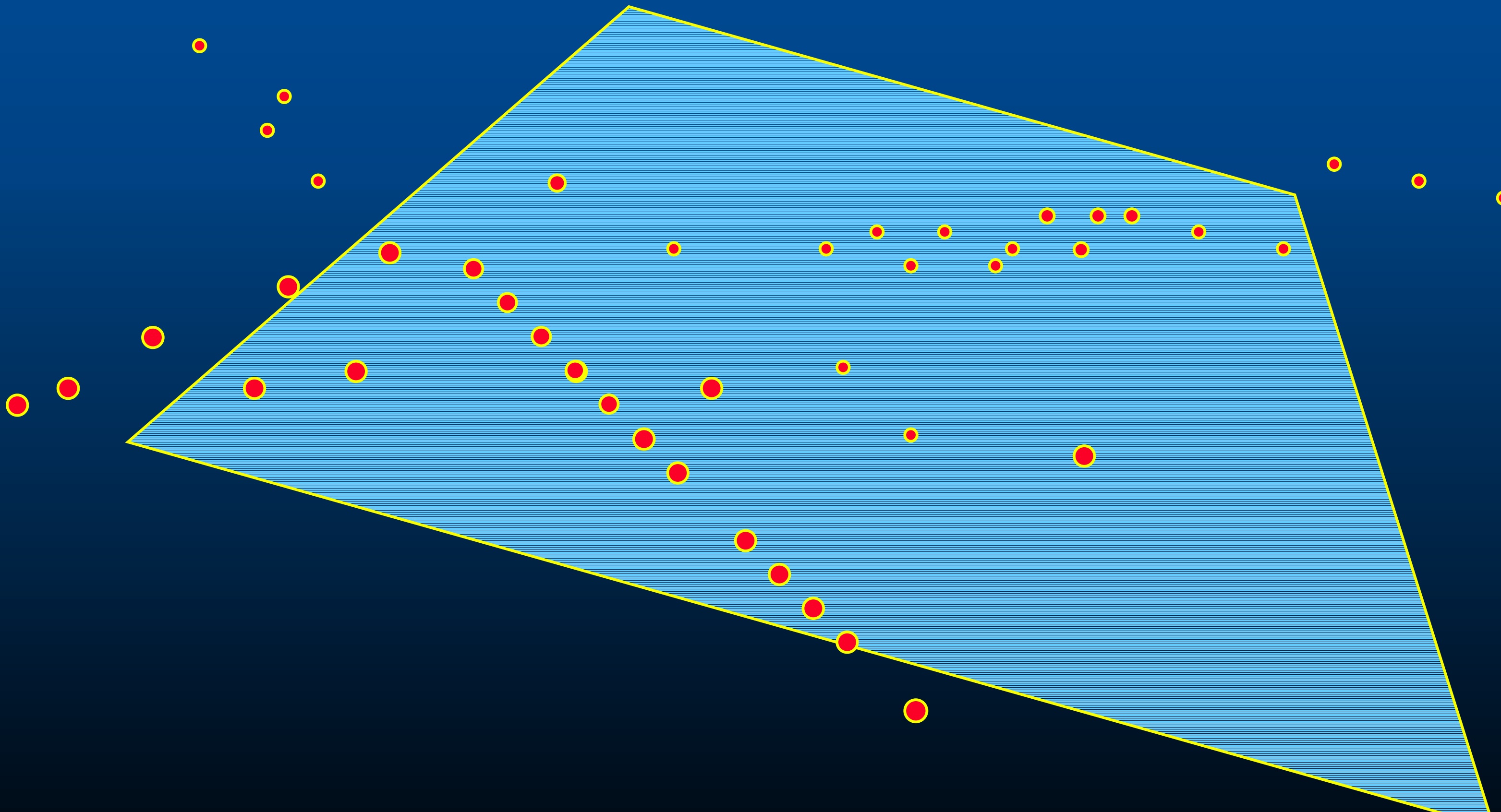
- 2 Views----Consider Image coordinates.

	Dimension 3	Dimension 2
Bilinear	**F** Matrix	Homography **H**
Linear	Affine **F** Matrix	Affinity **A**

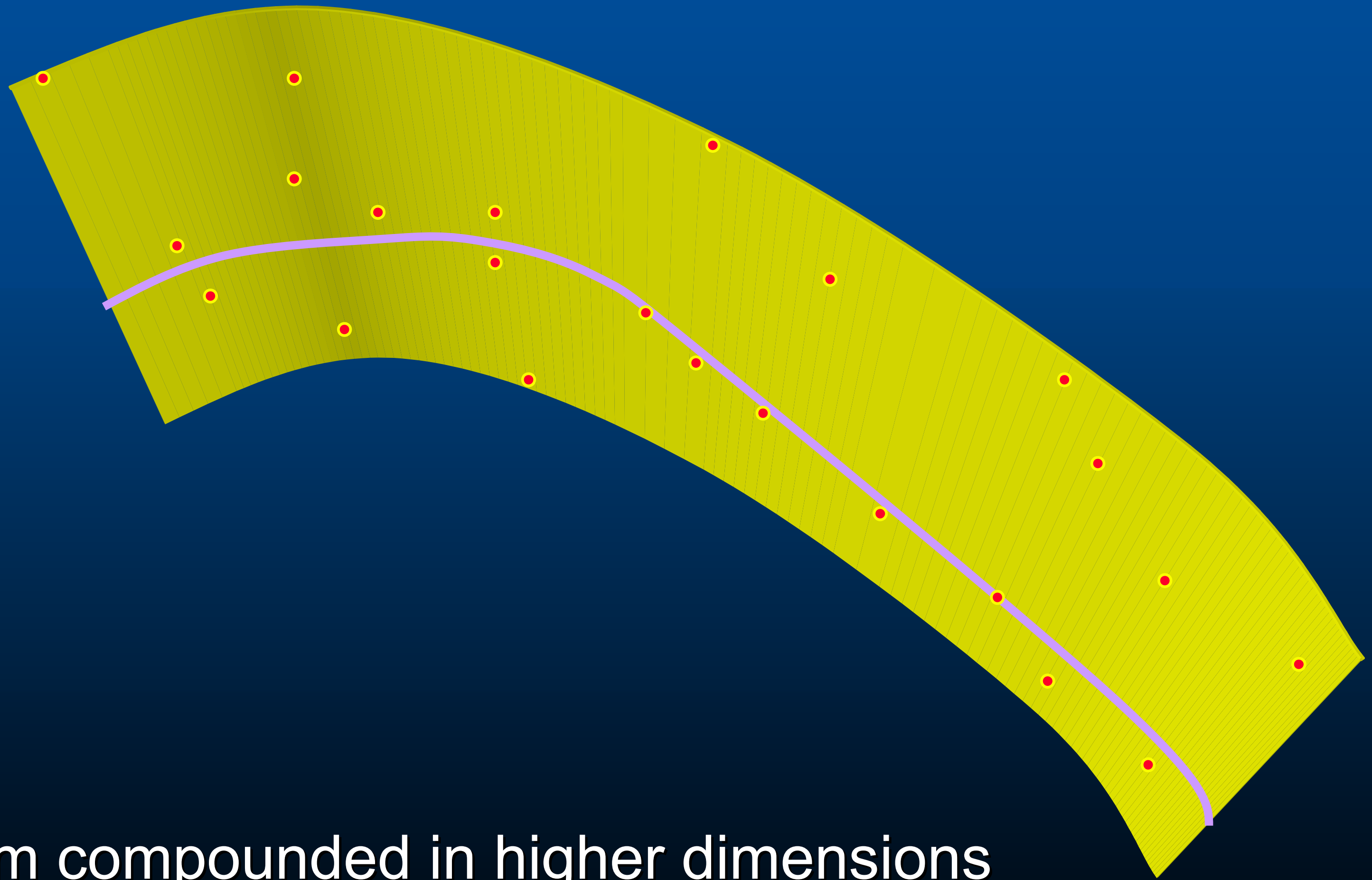
- (non generic: quadratic transformations)

Generic Problem

- Determine the degree and dimension of
- a manifold in d dimensional space.



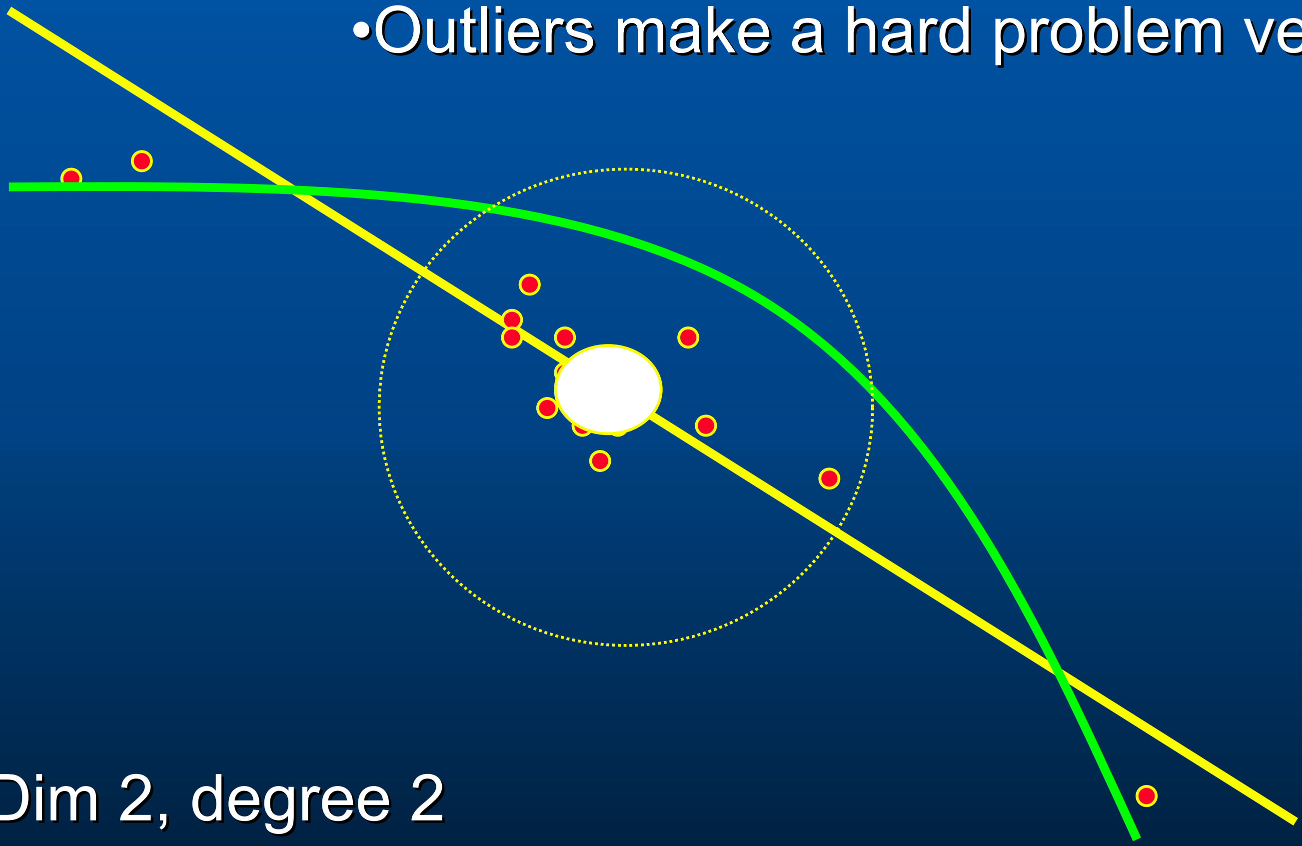
Examples



Problem compounded in higher dimensions

Robust Model Selection

- Outliers make a hard problem very hard!



rve Dim 2, degree 2

ne Dim 1, degree 1

oint Dim 0, degree 1

Error in Variable Model (EVM)

- Noise on points is (possibly a robust mixture):

$$\underline{x}_i^j = x_i^j + \epsilon_{ij}$$

$$\underline{y}_i^j = y_i^j + \gamma_{ij} \qquad j = 1, 2, \text{ and } i = 1 \ldots n$$

- Where points lie on a manifold defined by q implicit relations (if polynomial this is also a variety):

$$g_q(x_i^1, y_i^1, x_i^2, y_i^2, \boldsymbol{\theta}) = 0 \qquad i = 1 \ldots n, \text{ and } q = 1 \ldots Q.$$

Parameters:

on-Latent Variables

parameters of manifold (i.e. 7 for **F**)

atent Variables

Form of the Non-robust & Robust likelihood L

$$\Pr(\mathcal{D}|\boldsymbol{\theta},\mathcal{M},\mathcal{I}) = \Pr(\mathcal{D}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\mathcal{M},\mathcal{I}) = \Pr(\mathcal{D}|\boldsymbol{\beta},\boldsymbol{\gamma},\mathcal{M},\mathcal{I})$$

$$\Pr(\mathcal{D}|\boldsymbol{\beta},\mathcal{M},\mathcal{I}) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{nD} \prod_{i=1\ldots n} \exp - \frac{\sum_{j=1,2}(\hat{x}_i^j - x_i^j)^2 + (\hat{y}_i^j - y_i^j)^2)}{(2\sigma^2)},$$

$$\Pr\left(\mathcal{D}|\boldsymbol{\beta},\boldsymbol{\gamma},\mathcal{M},\mathcal{I}\right) = \prod_{i=1\ldots n} \psi(e_i) = \prod_{i=1\ldots n} \left(\gamma_i \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^D \exp\left(-\frac{e_i^2}{2\sigma^2}\right) + (1-\gamma_i)\frac{1}{v}\right)$$

Note

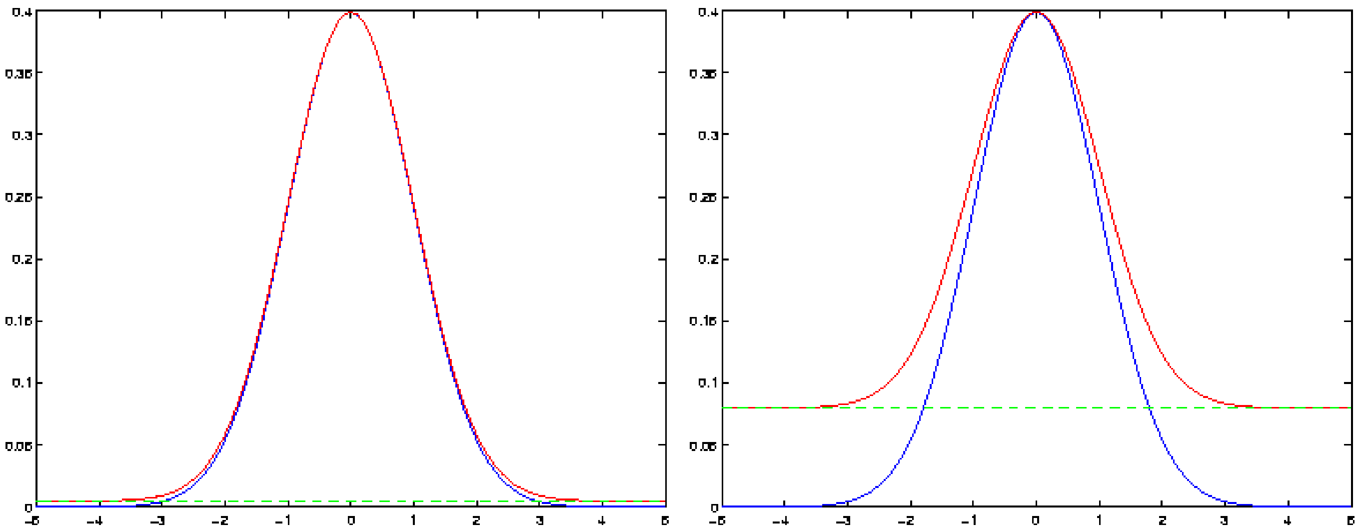
- The error distribution is rotationally symmetric in Gaussian and uniform case.

The manifold of F is locally flat (i.e. for small fields of view affine F is fine).

A possible approximation to robust function: maximize or marginalize?

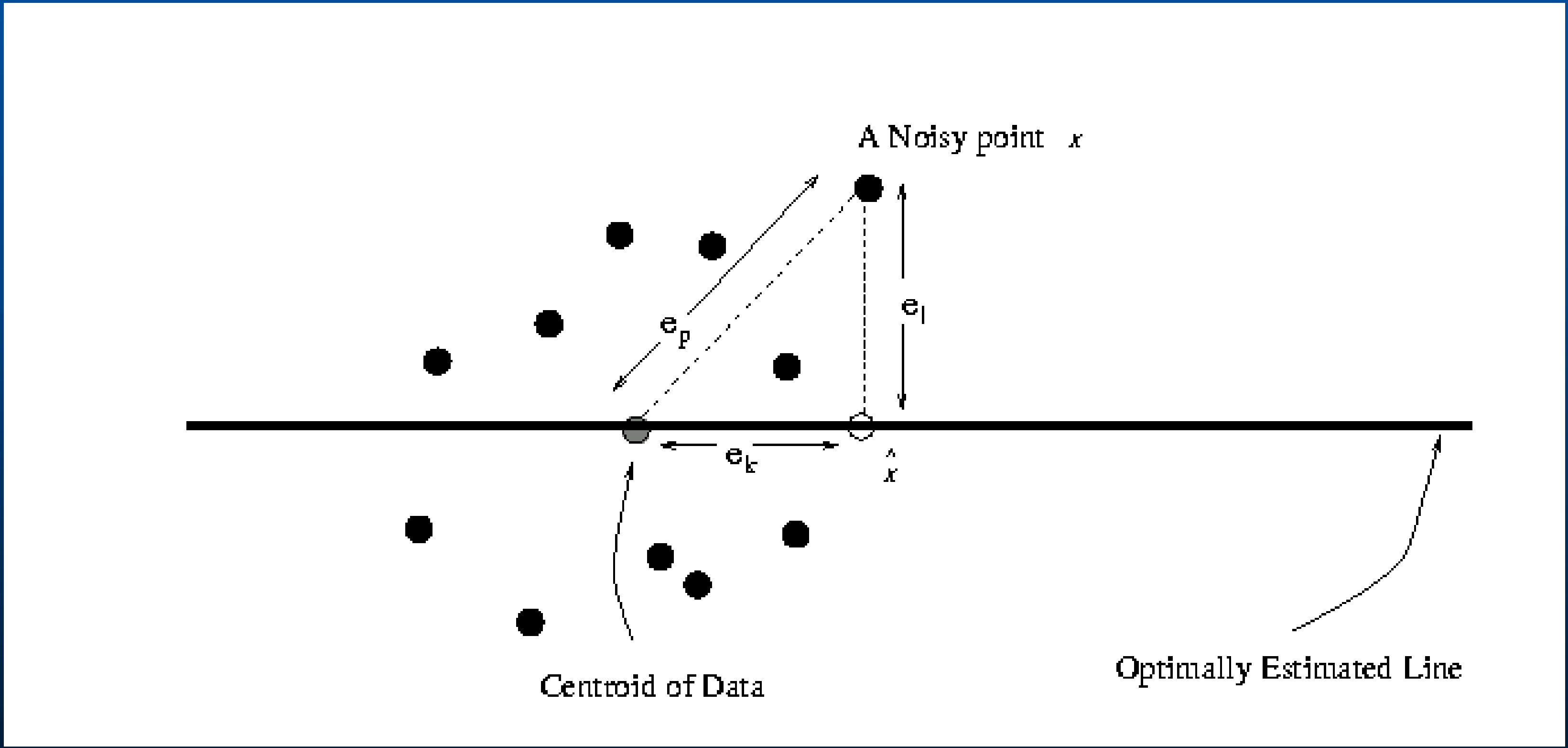
$$\rho_2\left(\frac{e^2}{\sigma^2}\right) = \begin{cases} \frac{e^2}{\sigma^2} & \frac{e^2}{\sigma^2} < T \\ T & \frac{e^2}{\sigma^2} \geq T\,. \end{cases}$$

- Or could marginalize over γ
- (or EM) [Torr 97]



- Red-mixture, green-uniform, blue-Gaussian.

Total least squares



Number of Parameters in Model

- The number of parameters in the system is
- typically: $k = p + n d$

- p number of parameters to define manifold
- d the dimension (2 or 3 for image sequences)
- n the number of features (data)

Summary of some two view relations

iii

Relation, $\mathcal{R}$	c	k	d	Q	Constraints, $g_q(x, y, x', y'; \boldsymbol{\theta}) = 0$	Parameters, $\boldsymbol{\theta}$
General	7	7	3	1	$\mathbf{x}^{2\top}\mathbf{F}\mathbf{x}^1 = 0$	$\mathbf{F} = \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix}$
Affine $\mathbf{F}_A$	4	4	3	1	$\mathbf{x}^{2\top}\mathbf{F}_A\mathbf{x}^1 = 0$	$\mathbf{F}_A = \begin{bmatrix} 0 & 0 & f_3 \\ 0 & 0 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix}$
Homography	4	8	2	2	$\mathbf{x}^2 = \mathbf{H}\mathbf{x}^1$	$\mathbf{H} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix}$
Affinity	3	6	2	2	$\mathbf{x}^2 = \mathbf{H}_A\mathbf{x}^1$	$\mathbf{H}_A = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ 0 & 0 & a_7 \end{bmatrix}$

TABLE 1. **Two View Relations;** *A description of the reduced models that are fitted to degenerate sets of correspondences. c is the minimum number of correspondences needed in a sample to estimate the constraint. k is the number of parameters in the relation; d is the dimension of the constraint, Q is the number of independent constraints $g_k()$ on the image coordinates.*

Many more possible

- No Motion
- Translation
- Quadratic Transform
- Multi view, 1-2,& 2-3 F but 1-3 H (Pollefreys)

Classical Model Selection: Hypothesis Testing

- Likelihood Ratio test:

$$\lambda(\mathcal{D}) = 2\log\left(\frac{\Pr(\mathcal{D}|\hat{\theta}_1\mathcal{M}_1))}{\Pr(\mathcal{D}|\hat{\theta}_2\mathcal{M}_2))}\right) = 2(L_1 - L_2)$$

- Follows chi square distribution:

$$\lambda(\mathcal{D}) = 2(L_1 - L_2) < \chi^2(\alpha, p_1 - p_2)$$

Problems with Hypothesis Testing

- Hard to apply to non nested models or when there are multiple models to choose between.

- This lead to a host of penalized likelihood methods being proposed…

Some penalty model selection schemes

- It all started with Mallow's:

Author	Criterion
Mallows' [12] C_p	$-2\log L - n + 2k$
Akaike's [1] AIC	$-2\log L + 2k$
Schwarz [15]	$-2\log L + 2k\log n$
Schwarz [15] KC	$-2\log L - \log p_r + \log\lvert\Sigma\rvert + k\log n$
Rissanen's [14] SSD	$-2\log L + k\log\frac{n+2}{24} + 2\log(k+1)$
Rissanen's [14] MDL	$-2\log L + \frac{1}{2}k\log n$
Bozdogan's [2] CAIC	$-2\log L + k(\log(n)+1)$
Bozdogan's [2] CAICF	$-2\log L + k(\log(n)+2) + \log\lvert\mathbf{J}(\theta_k)\rvert$
Wallace's [20] MML2	$-2\log L - \log p_r + \frac{1}{2}(\log\lvert\mathbf{J}(\theta_k) + k)\rvert$

Notes:

- Model selection schemes are Max likelihood plus a penalty related to k.
- Mallows similar to AIC
- Kanatani used AIC

GRIC

- GRIC similar but exploits specific manifold structure:

$$\mathbf{GRIC} = -2L + \lambda_1 nd + \lambda_2 k$$

- Derivation of γ postponed.

Problem with AIC &c

- Tends to over fit, due to ‘magic number’ 2.
- Inconsistent when compared to chi squared tes

$$2(L_1 - L_2) < 2(p_1 - p_2)$$

Chance of over fit for

$$\chi^2(\alpha, p_1 - p_2)$$

$\lvert p_1 - p_2 \rvert$	1	2	3
α	0.156	0.135	0.111

Test of AIC

Estimated		Point Motion			
		General $\mathbf{F}$	Orthographic $\mathbf{F}_A$	Rotation $\mathbf{H}$	Affinity $\mathbf{F}_A$
Fundamental	$\hat{\mathbf{F}}$	**99**	12	0	0
Affine	$\hat{\mathbf{F}}_A$	1	**88**	0	0
Homography	$\hat{\mathbf{H}}$	0	0	**98**	15
Affinity	$\hat{\mathbf{H}}_A$	0	0	2	**85**

BLE 6. *Number of times each model selected over 100 trials, using* AIC *for each of the four motion type* *be seen that* AIC *tends to overfit the degree of the model.*

Minimum Description Length: MDL

Goal to find model that optimally compresses data
Approximation to stochastic complexity:

$$\mathrm{MDL} = -2L - \frac{p}{2}\log N$$

- Asymptotically over fits as N increases.
- Contrary to popular belief it is non-Bayesian.

Test of BIC/MDL, under fits.

Estimated		Point Motion			
		General $\mathbf{F}$	Orthographic $\mathbf{F}_A$	Rotation $\mathbf{H}$	Affinity $\mathbf{F}_A$
Fundamental	$\hat{\mathbf{F}}$	**69**	1	0	0
Affine	$\hat{\mathbf{F}}_A$	17	**73**	0	0
Homography	$\hat{\mathbf{H}}$	14	25	**96**	2
Affinity	$\hat{\mathbf{H}}_A$	0	1	4	**98**

TABLE 7. *Number of times each model selected over 100 trials, using* BIC *for each of the four motion type*

- BIC ('Bayesian' Information Criterion) is the same form as MDL and is not really Bayesian.

Difference Between MDL & Bayes

- MDL is an approximation to stochastic complexity which is uncomputable, a big proble for any theory, Bayesian solution is, in many cases computable.

- MDL=maximum compression
- Bayes = maximize utility.

[Peter Grunwald, Li & Vitanyi]

Bayesian Model Comparison

$$\Pr(\mathcal{M}_i|\mathcal{D},\mathcal{I}) = \frac{\Pr(\mathcal{D}|\mathcal{M}_i,\mathcal{I})\Pr(\mathcal{M}_i|\mathcal{I})}{\sum_{j=1}^{j=K}\Pr(\mathcal{D}|\mathcal{M}_j,\mathcal{I})\Pr(\mathcal{M}_j|\mathcal{I})}$$

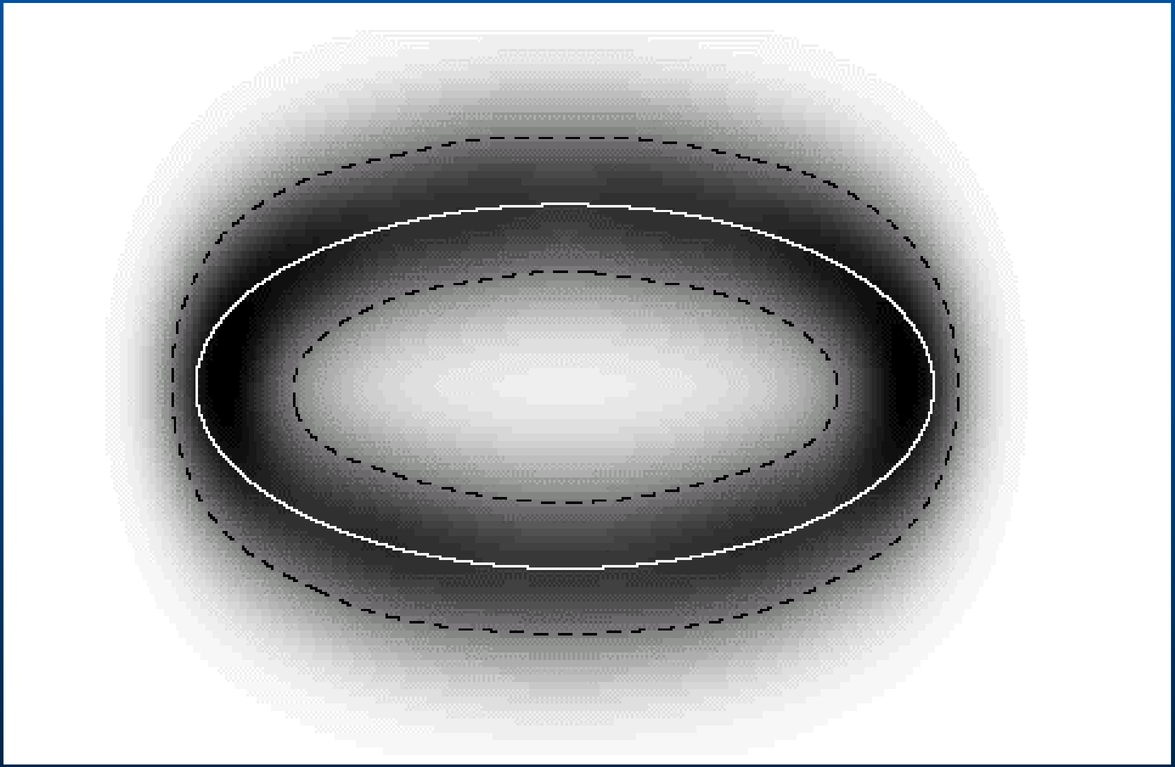
$$\begin{aligned}\Pr(\mathcal{D}|\mathcal{M}_j,\mathcal{I}) &= \int \Pr(\mathcal{D}|\mathcal{M}_j,\boldsymbol{\theta}_j,\mathcal{I})\Pr(\boldsymbol{\theta}_j|\mathcal{M}_j,\mathcal{I})d\boldsymbol{\theta}_j \\ \text{Evidence} &= \int \text{likelihood} \times \text{prior}\, d\boldsymbol{\theta}_j \quad ,\end{aligned}$$

A Horrible integral:

$$\Pr(\mathcal{D}|\mathcal{M}_j\mathcal{I}) = \int_{\boldsymbol{\alpha}}\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\gamma}}\prod_{i=1\ldots n}(\psi(e_i)\Pr(\boldsymbol{\beta}_i|\boldsymbol{\alpha},\mathcal{M},\mathcal{I}))\Pr(\boldsymbol{\gamma}|\boldsymbol{\alpha},\mathcal{M},\mathcal{I})\Pr(\boldsymbol{\alpha}|\mathcal{M},\mathcal{I})\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}$$

- Can we simplify it?
- Integrate out the β and γ.

Distribution of β given α

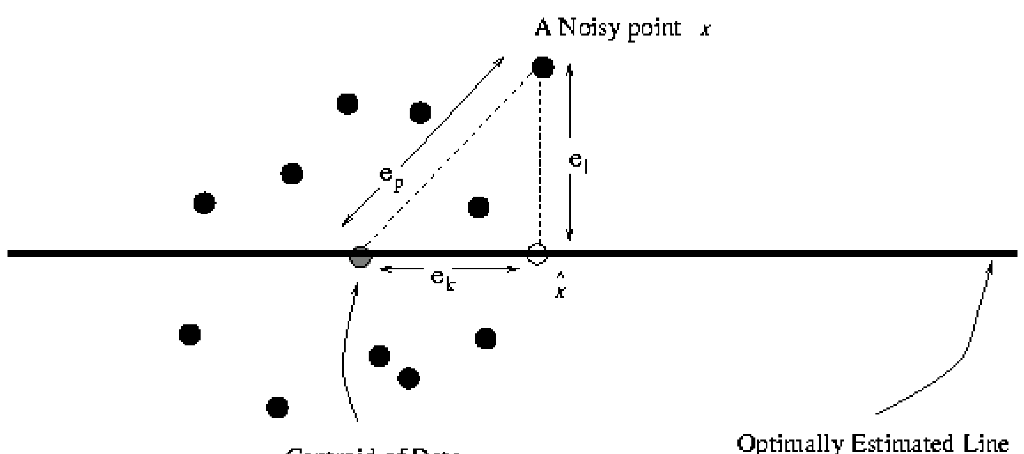


- Important observation: likelihood proportional to distribution of β on α.

Distribution of β given α

- **Assuming local flatness**, given α then β can be determined using the following identity:

$$\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^W \int \exp\left(-\frac{\mathbf{w}^\top \mathbf{w}}{2\sigma^2}\right) \partial \mathbf{w} = 1,$$



Distribution of α
β given α, assuming uniform distribution on manifold.

$$\left(\gamma_i \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^D \exp\left(-\frac{e_\perp^2 + e_\parallel^2}{2\sigma^2}\right)\right) \frac{1}{c}\partial\boldsymbol{\beta}_i \approx \gamma_i \left(\frac{(\sqrt{2\pi\sigma^2})^{d-D}}{c} \exp\left(-\frac{e_\perp^2}{2\sigma^2}\right)\right)$$

- C is the area of the manifold (note the manifold is of finite extent in the joint image space defined by the image boundaries).

$$c = \int_{\Re^d} \partial\boldsymbol{\beta}_i \approx \mathcal{O}(L^d)$$

Distribution of α
β given α, for robust part of mixture.

$$\int \left((1-\gamma_i)\frac{1}{v} \right) \frac{1}{c} \partial \beta_i = (1-\gamma_i)\frac{1}{v}.$$

- Thus the integral is still a mixture.

$$\int \psi(e_i) \Pr(\beta_i | \alpha, \mathcal{M}, \mathcal{I}) \partial \beta_i = \gamma_i \left(\frac{(\sqrt{2\pi\sigma^2})^{d-D}}{c} \exp\left(-\frac{e_\perp^2}{2\sigma^2} \right) \right) + (1-\gamma_i)\frac{1}{v},$$

λ_1 for Gaussian:

$$\mathrm{GRIC} = -2L + \lambda_1 nd + \lambda_2 k$$

- From the analysis above, for all α:

$$\lambda_1 = \log\left(\frac{U^2}{2\pi\sigma}\right),$$

- Assuming a uniform distribution on manifold then U = L.

f some other distribution on manifold (i.e. robust) t

Robust Evidence is now:

- Taking expectations over β and γ_i.

$$\mathcal{D}|\mathcal{M}_j,\mathcal{I}) = \int_{\boldsymbol{\alpha}} \prod_{i=1\dots n} \left(\gamma\left(\frac{(\sqrt{2\pi\sigma^2})^{d-D}}{c}\exp\left(-\frac{e_i^2(\boldsymbol{\alpha})}{2\sigma^2}\right)\right) + (1-\gamma)\frac{1}{v}\right)\Pr(\boldsymbol{\alpha}|\mathcal{M},\mathcal{I})$$

$\hat{\boldsymbol{\alpha}}$ be the MAP estimate, and define

$$L_{\mathrm{MAP}} = \sum_{i=1\dots n} \log\left(\gamma\left(\frac{(\sqrt{2\pi\sigma^2})^{d-D}}{c}\exp\left(-\frac{e_i^2(\hat{\boldsymbol{\alpha}})}{2\sigma^2}\right)\right) + (1-\gamma$$

hen as $\int_{\boldsymbol{\alpha}} \partial\boldsymbol{\alpha} = 1$, the minimum value of $-\log\left(\Pr(\mathcal{D}|\mathcal{M}_j,\mathcal{I})\right) = -L_{\mathrm{M}}$

Intuition

- GRIC works as long as
  - The error distribution is rotationally symmetric.
  - The manifold has mostly low curvature.

- Then we can integrate out the latent variables whatever the value of α.

Laplace's Approximation

- As the number of observations increases α becomes normal with covariance approximated by the inverse Hessian Λ

$$\log(\Pr(\mathcal{D}|\mathcal{M},\mathcal{I})) \approx \log(\Pr(\mathcal{D}|\mathcal{M},\boldsymbol{\theta},\mathcal{I})) + \frac{p}{2}\log 2\pi + \frac{1}{2}\log|\Lambda| + \log(\Pr(\hat{\boldsymbol{\theta}}|\mathcal{M},\mathcal{I}))$$

- Note problem when model is unidentifiable.

GRIC approximation

- Use BIC approximation for α

$$\mathrm{GRIC} = -2L_{\mathrm{MAP}} + k \log n,$$

$$\lambda_2 = k \log(n)$$

- Approximation for robust case:

$$\mathrm{GRIC} = \sum_i \rho_2\left(\frac{e_i^2}{\sigma^2}\right) + \lambda_1 n d + \lambda_2 k + \mathrm{constant}$$

$$\rho_2\left(\frac{e^2}{\sigma^2}\right) = \begin{cases} \frac{e^2}{\sigma^2} & \frac{e^2}{\sigma^2} < T \\ \mathrm{constant} & \frac{e^2}{\sigma^2} \geq T \end{cases}.$$

Bounds

- Thus we have an absolute upper bound and approximate lower bound on the evidence.
- Experiments reveal the solution is reasonably invariant to the choice of lower bound (GRIC)
- Compare GRIC’s if close not enough evidence to distinguish otherwise pick lowest.

Invariant prior on α

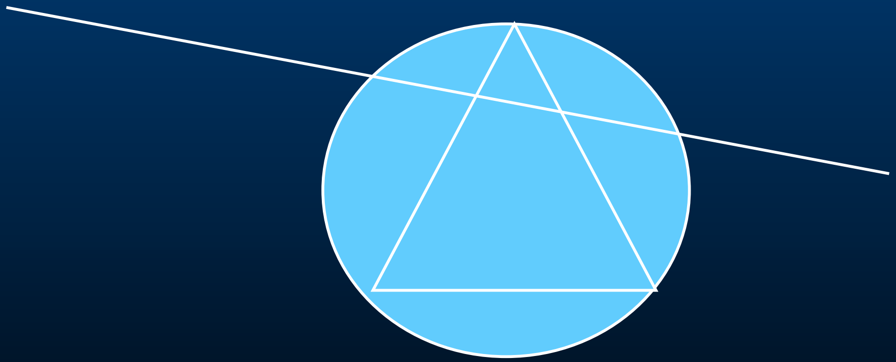
- Use result from Stochastic Geometry [Kendall, Santalo]; e.g. for a line:

$$ux + vy + 1 = 0$$

$$\Pr(\boldsymbol{\alpha}|\mathcal{M},\mathcal{I}) = \Pr(u,v|\mathcal{M},\mathcal{I}) \propto \frac{1}{(u^2+v^2)^{\frac{3}{2}}}$$

Related to Betrand's Paradox

- Throw straws at a circle, what is probability that chord will have length greater than the side of enscribed equilateral triangle.

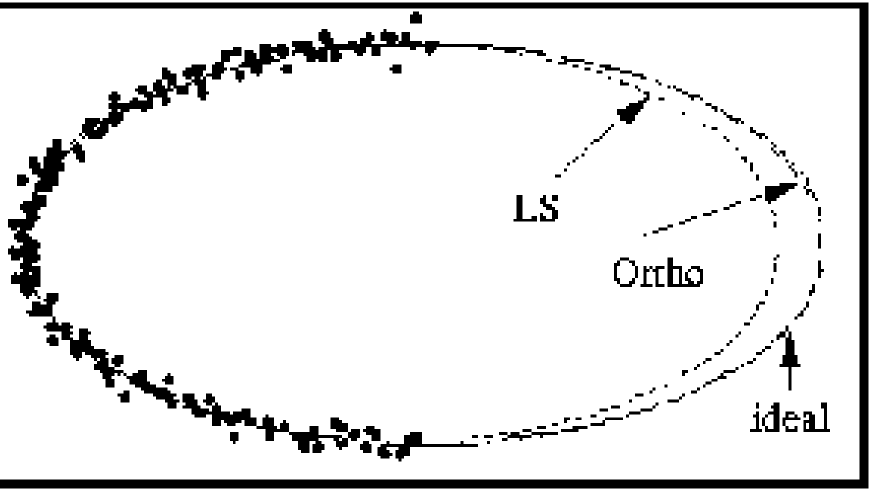


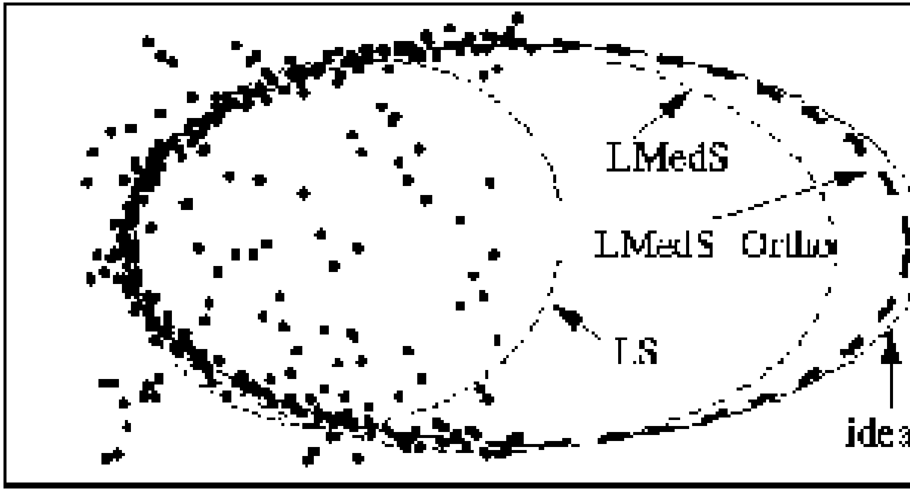
- Depends on distribution of lines, which sho not depend on coordinate syste

Robust estimator: MLESAC

- Take minimal number of matches to estimate two view relation.
- maximize posterior (MLESAC)
- Provides better results to RANSAC.

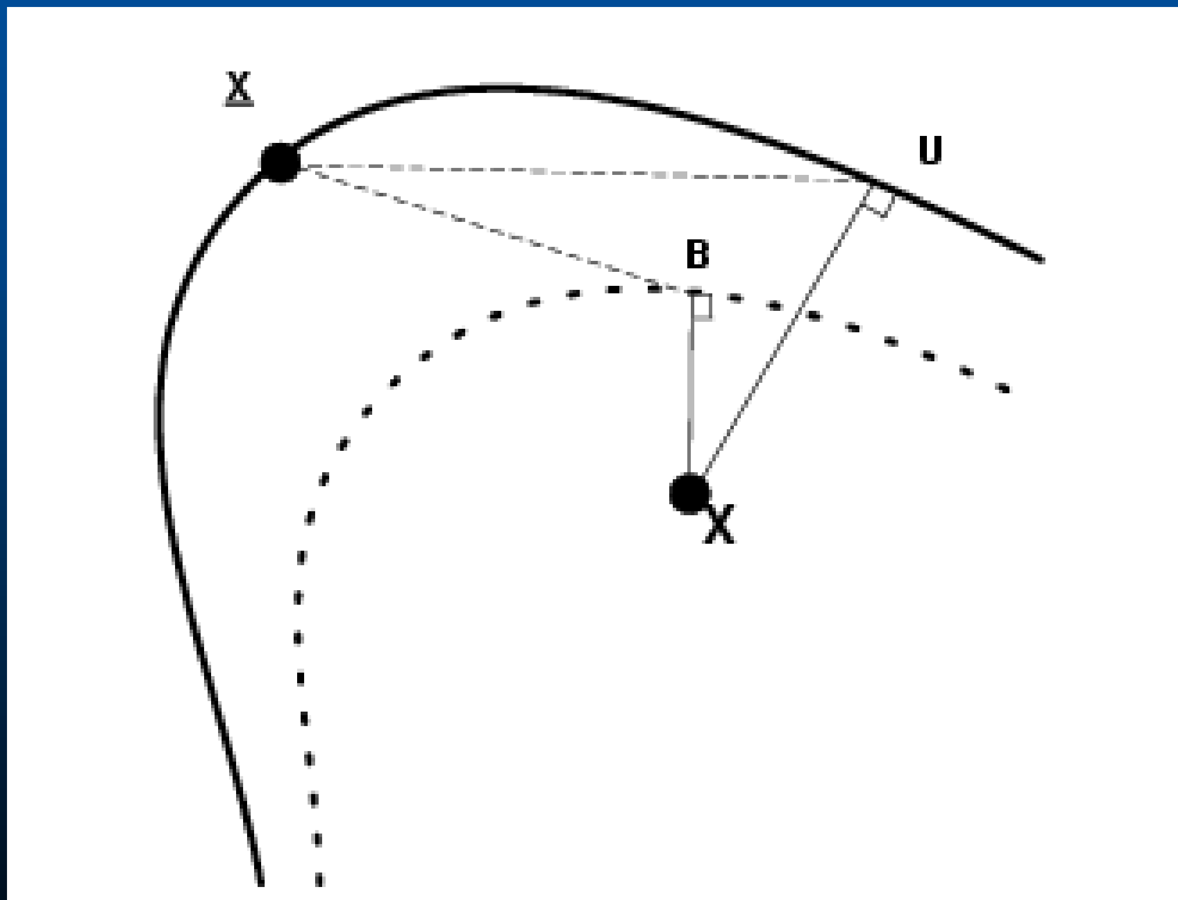
Maximize over α; maximize over α and β?



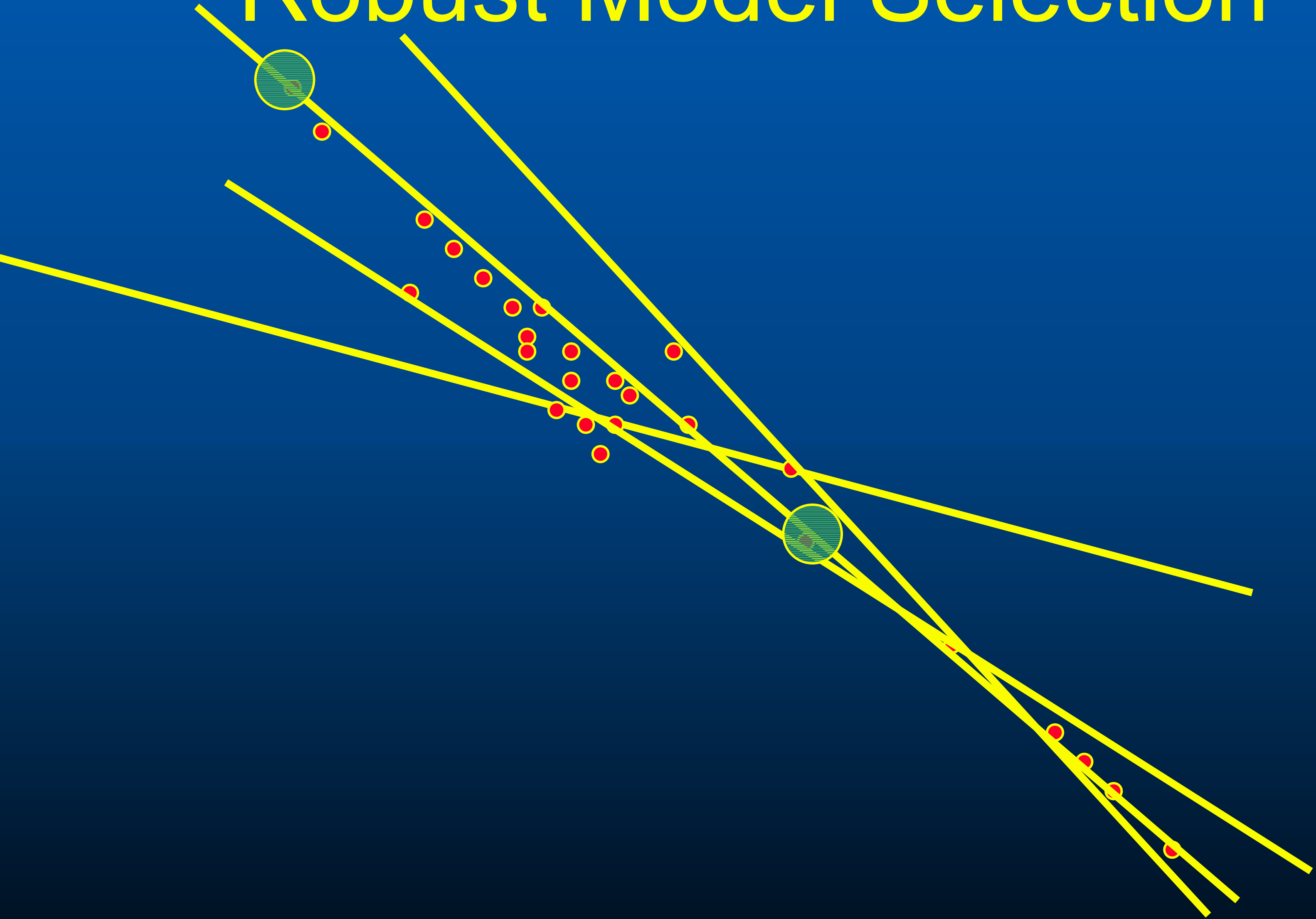


- If we don't marginalize β we get a biased (higher curvature) fit to α.
- However we get less accurate β.

Maximize over α and β



Robust Model Selection



Testing-see Torr 2002

- Testing is on going, code Matlab Tool kit for this and other SF will be online in May (to coincide with ECCV).
- Methodology: Generate synthetic data with varying noise and see whether correct model selected.
- Actually experiments reveal that the results of the model selection are invariant over a wide range of λ's; indicating in general the choice is not crucial.

Results of GRIC

Estimated		Point Motion			
		General $\mathbf{F}$	Orthographic $\mathbf{F}_A$	Rotation $\mathbf{H}$	Affinity $\mathbf{F}_A$
Fundamental	$\hat{\mathbf{F}}$	**97**	1	1	0
Affine	$\hat{\mathbf{F}}_A$	1	**95**	0	2
Homography	$\hat{\mathbf{H}}$	2	2	**99**	2
Affinity	$\hat{\mathbf{H}}_A$	0	2	0	**96**

TABLE 9. *Number of times each model selected over 100 trials, with outliers, using robust* GRIC *for each of the four motion types.*

Real Image Results

Image 1 | Image 2 with matches

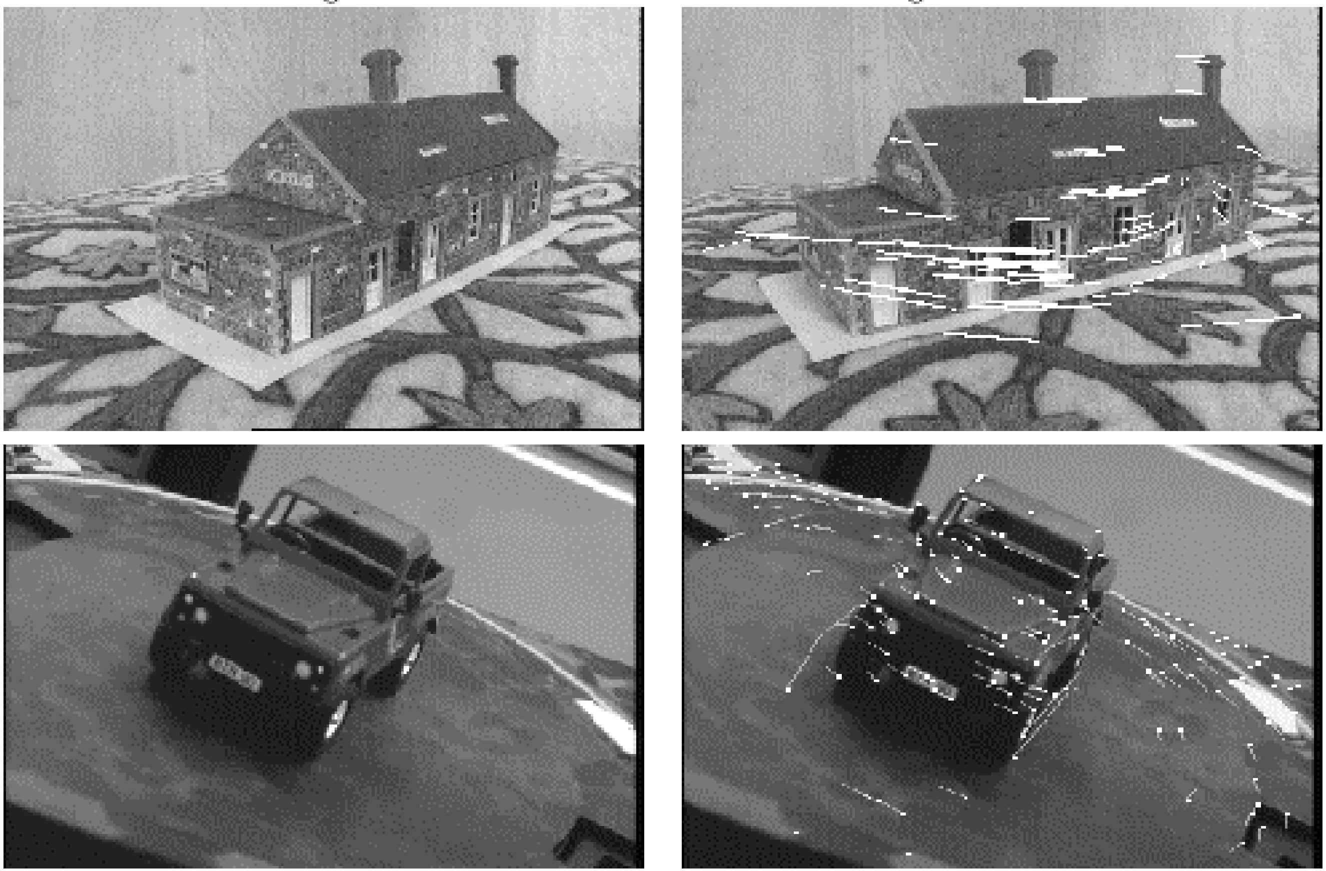


FIGURE 3. *Row 1 Indoor sequence, camera translating and rotating to fixate on the house. Model selected is* **F**. *Row 2 two views of a buggy rotating on a table. Model selected is* $\mathbf{F}_A$. *With disparity vectors for features*

Real Image

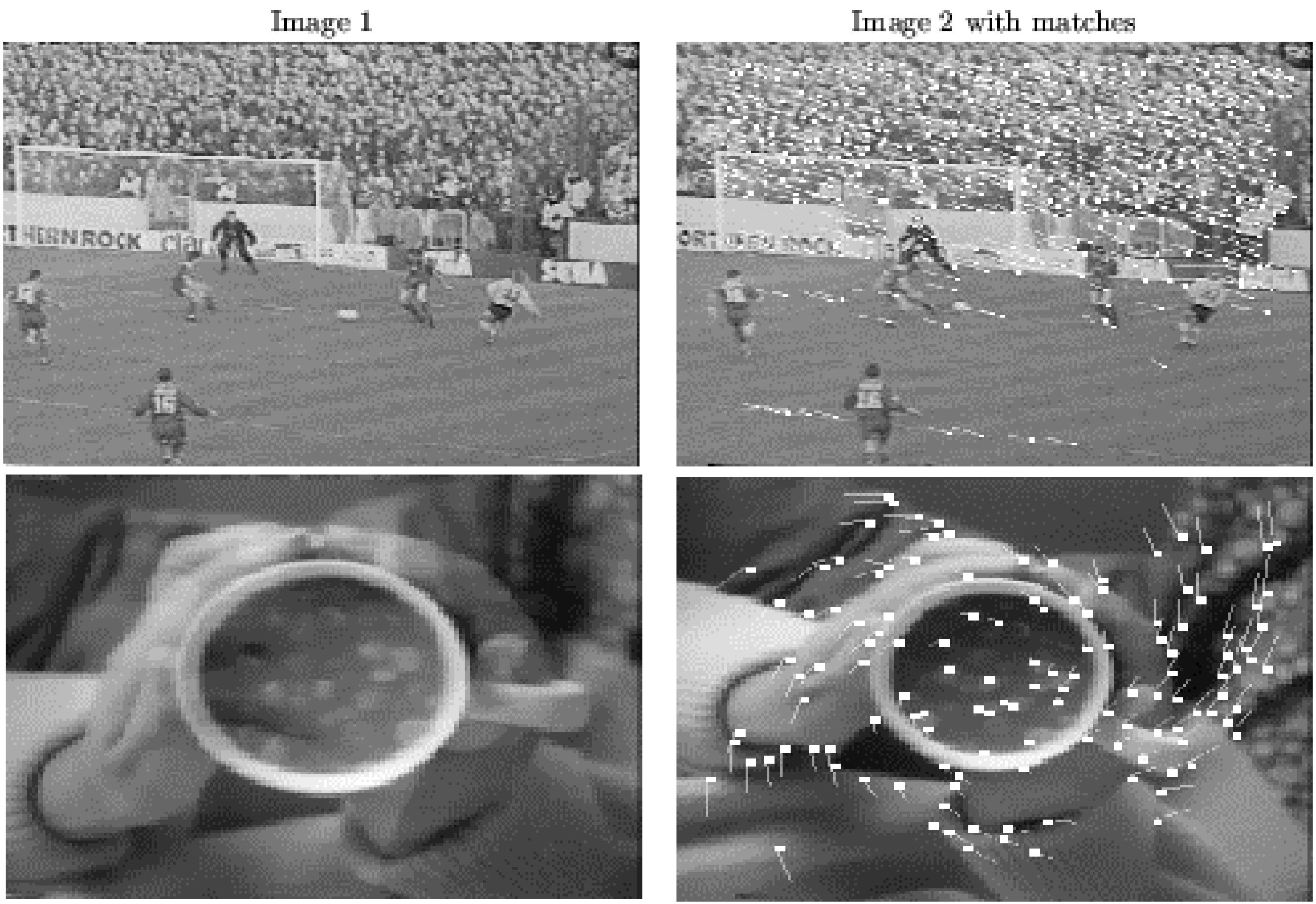


FIGURE 4. *Row 1 Football matches, camera panning. Model selected is* $\mathbf{H}$. *Row 2 Cup, camera zooming and cyclo-rotating. Model selected is* $\mathbf{H}_A$. *With disparity vectors for features superimposed.*

Conclusion

- Presented a Bayesian analysis of model selection problems.
- The analysis provides rough approximations to what the λ's should be
- The results are reasonable even if λ's are approximated.

END

- Matlab Code and paper available:
- Matching, est F, sfm, segmentation.
- http://www.research.microsoft.com
- philtorr@microsoft.com